2008/60

■

Barrier subgradient method

Yu. Nesterov

Barrier subgradient method

Yu. NESTEROV[1]

October 2008

Abstract

In this paper we develop a new primal-dual subgradient method for nonsmooth convex optimization problems. This scheme is based on a self-concordant barrier for the basic feasible set. It is suitable for finding approximate solutions with certain *relative accuracy*. We discuss some applications of this technique including fractional covering problem, maximal concurrent flow problem, semidefinite relaxations and nonlinear online optimization.

Keywords: convex optimization, subgradient methods, non-smooth optimization, minimax problems, saddle points, variational inequalities, stochastic optimization, black-box methods, lower complexity bounds.

1 Introduction

Motivation. During the last two decades we can see an increasing interest to the Structural Convex Optimization. This direction of research, which can be seen as an alternative to the black-box computational model [6], can be traced from the theory of self-concordant functions [16], through the series of results on fully polynomial approximation schemes [2, 3, 4, 18], and up to the recent improvements related to the smoothing technique [1, 9, 10, 12, 14] and applications of the extra-gradient methods for smooth variational inequalities [5, 11]. We should mention also recently published results [13] for constructing primal-dual subgradient methods on the basis of available information on the problem's structure. In order to see a clear picture of these developments and discover the missing pieces, let us present some classification of the above results.

In Nonlinear Optimization, the performance of numerical methods is closely related to our abilities to execute some auxiliary computations related to the convex sets involved in the problem's formulation. Usually, the optimization methods assume feasibility of one of the following operations:

L: Maximization of a linear function $\langle c, x \rangle$ over a convex set Q.

S: Maximization of the difference $\langle c, x \rangle - d(x)$ in $x \in Q$, where d is a strongly convex prox-function of the set Q.

B: Computation of the value and first two derivatives of some self-concordant barrier at the interior points of convex set Q.

Note that in Structural Optimization we always can consider the optimization problem posed in a primal-dual form. The most important example of such a representation is a bilinear saddle point formulation:

$$\min_{x \in Q_p} \max_{w \in Q_d} \langle Ax, w \rangle. \tag{1.1}$$

where Q_p and Q_d are closed convex sets in corresponding spaces and A is a linear operator. Since the structure of the primal and dual sets may be of different complexity, we have six possible combinations of the above mentioned auxiliary operations. Let us present the known results on their complexity.

- $\mathbf{L_p} \bigcap \mathbf{L_d}$: to the best of our knowledge, no complexity results are obtained for this combination.

- $\mathbf{S_p} \bigcap \mathbf{S_d}$: this case is treated by the smoothing technique [9, 10]. An ϵ-solution of the problem (1.1) can be obtained in

$$O\left(\tfrac{1}{\epsilon} \cdot \|A\| \cdot [D_1 D_2]^{1/2} \right)$$

  gradient steps, where D_1 and D_2 are the *sizes* of the primal and dual sets, and the norm $\|A\|$ is defined by the norms of the primal and dual spaces.

- $\mathbf{B_p} \bigcap \mathbf{B_d}$. In this situation, Interior-Point Methods [16] provide an ϵ-solution of the problem (1.1) in

$$O\left(\sqrt{\nu} \cdot \ln \tfrac{\nu}{\epsilon} \right)$$

  Newton steps, where ν is the parameter of a self-concordant barrier for primal-dual feasible set $Q_p \times Q_d$.

2

- $\mathbf{S_p} \bigcap \mathbf{L_d}$. This case is similar to the standard Black-Box Nonsmooth Minimization. Primal-dual subgradient methods [13] provide an ϵ-solution to (1.1) in

$$O\left(\tfrac{1}{\epsilon^2} \cdot \|A\|^2 \cdot D_1 \cdot D_2 \right)$$

  gradient steps.

- $\mathbf{B_p} \bigcap \mathbf{S_d}$. Complexity of this combination is not known yet.

- The last variant $\mathbf{B_p} \bigcap \mathbf{L_d}$ is studied in the present paper. From the view point of Black-Box Optimization, it corresponds to a problem of minimizing nonsmooth convex function over a feasible set endowed by a self-concordant barrier.

Contents. The paper is organized as follows. In Section 2 we study the smoothing of support function of convex set by a self-concordant barrier. The corresponding *barrier subgradient method* (BSM) is described in Section 3. As it is shown in Section 4, this method can be naturally applied to the problem of maximizing a non-negative concave function. In this case we can obtain the complexity bounds for computing an approximate solution in *relative scale*. In Section 5 we consider several applications: fractional covering problem, maximal concurrent flow problem, minimax problem with nonnegative components, and semidefinite relaxation of boolean quadratic problem. Finally, in Section 6 we consider the applications of BSM to *Online Optimization*. We argue that this approach can be seen as a reasonable alternative to the single-stage Stochastic Optimization. In Appendix we discuss the complexity of barrier projection onto a standard simplex.

Notation. In what follows, we denote by E (or other capital letter) a finite dimensional linear vector space. The dual space (the space of linear functions on E) is denoted by E^*. We denote the value of function $s \in E^*$ on $x \in E$ by $\langle s, x \rangle$. This notation is used for all linear spaces employed in the paper. Thus, the actual meaning of $\langle \cdot, \cdot \rangle$ can be understood from the spaces of the arguments. For example, for the coordinate space $E = R^n$ with $E^* = R^n$, this notation has the following sense:

$$\langle s, x \rangle \;=\; \sum_{i=1}^{n} s^{(i)} x^{(i)}, \quad x, s \in R^n.$$

For a linear operator $A : E \to H^*$ we denote by $A^* : H \to E^*$ the adjoint operator:

$$\langle Ax, y \rangle \;=\; \langle A^* y, x \rangle, \quad x \in E, \; y \in H.$$

For a positive-semidefinte self-adjoint linear operator $B : E \to E^*$ we use notation $B \succeq 0$.

For a concave function $f(x)$ we denote by $\nabla f(x)$ one of its subgradients at x:

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle, \quad y, x \in \mathrm{dom}\, f.$$

For function of two vector variables $\Psi(u, x)$ notation $\nabla_2 \Psi(u, x)$ is used for its subgradient with respect to the second argument.

2 Smoothing by self-concordant barrier

Let $Q \subset E$ be a closed convex set containing no straight line. We assume Q be endowed with a ν-self-concordant barrier F. Consider another closed convex set $\hat{P} \subseteq E$. The set

of our main interest is
$$P = \hat{P} \bigcap Q,$$
which we assume to be bounded. Denote by x_0 its *constrained analytic center*:
$$x_0 = \arg\min_{x \in P_0} F(x) \in P_0 \stackrel{\text{def}}{=} \hat{P} \bigcap \text{int}\, Q \subseteq P. \tag{2.1}$$

Thus, $F(x) \geq F(x_0)$ for all $x \in P$. Since Q contains no straight line, x_0 is well defined (e.g. Theorem 4.1.3 [8]).

For the set P, we introduce the following smooth approximation of its support function:
$$U_\beta(s) = \max_{x \in \hat{P}} \{\langle s, x - x_0 \rangle - \beta[F(x) - F(x_0)]\}, \quad s \in E^*, \tag{2.2}$$

where $\beta > 0$ is a smoothing parameter. Denote by $u_\beta^*(s)$ the unique solution of the maximization problem (2.2). Then
$$\nabla U_\beta(s) = u_\beta^*(s) - x_0, \quad s \in E^*. \tag{2.3}$$

For any $x \in \text{int}\, Q$, consider the following local norms:
$$\|h\|_x = \langle \nabla^2 F(x)h, h \rangle^{1/2}, \qquad h \in E,$$

$$\|s\|_x^* = \langle s, [\nabla^2 F(x)]^{-1} s \rangle^{1/2}, \quad s \in E^*.$$

Then, we can guarantee the following level of smoothness of function $U_\beta(\cdot)$.

Lemma 1 *Let $\beta > 0$, $s \in E^*$ and $x = u_\beta^*(s)$. Then for any $g \in E^*$ with $\|g\|_x^* < \beta$ we have*
$$U_\beta(s + g) \leq U_\beta(s) + \langle g, \nabla U_\beta(s) \rangle + \beta \omega_*(\tfrac{1}{\beta}\|g\|_x^*), \tag{2.4}$$
where $\omega_*(\tau) = -\tau - \ln(1-\tau) \leq \frac{\tau^2}{2(1-\tau)}$.

Proof:
In view of definition (2.2), for any $y \in P_0$ we have
$$\langle s - \beta\nabla F(x), y - x \rangle \leq 0. \tag{2.5}$$

Moreover, since F is a self-concordant function, at any point $y \in \text{int}\, Q$
$$F(y) \geq F(x) + \langle \nabla F(x), y - x \rangle + \omega(\|y - x\|_x), \tag{2.6}$$
where $\omega(t) = t - \ln(1 + t)$ (see Theorem 4.1.7 [8]). Hence,

$U_\beta(s + g) - U_\beta(s) - \langle g, \nabla U_\beta(s) \rangle$

$\stackrel{(2.3)}{=} \max_{y \in P_0} \{\langle s + g, y - x_0 \rangle - \beta[F(y) - F(x_0)] - \langle s + g, x - x_0 \rangle + \beta[F(x) - F(x_0)]\}$

$= \max_{y \in P_0} \{\langle s + g, y - x \rangle - \beta[F(y) - F(x)]\}$

$\stackrel{(2.5)}{\leq} \max_{y \in P_0} \{\langle g, y - x \rangle + \beta[\langle \nabla F(x), y - x \rangle - F(y) + F(x)]\}$

$\stackrel{(2.6)}{\leq} \max_{y \in P_0} \{\langle g, y - x \rangle - \beta\omega(\|y - x\|_x)\} \leq \sup_{\tau \geq 0} \{\tau\|g\|_x^* - \beta\omega(\tau)\}.$

4

If $\|g\|_x^* < \beta$, then the supremum in the right-hand side is equal to $\beta\omega_*(\frac{1}{\beta}\|g\|_x^*)$. $\quad\square$

Consider now a linear function $l(x)$, $x \in P$. Denote $l_0 = l(x_0)$, $l^* = \max\limits_{x \in P} l(x)$, and

$$l^*(\beta) \;=\; \max_{y \in P_0}\{l(y) - \beta[F(y) - F(x_0)]\} \;\geq\; l_0 \tag{2.7}$$

for $\beta \geq 0$.

Lemma 2 *For any $\beta > 0$ we have*

$$l^*(\beta) \;\leq\; l^* \;\leq\; l^*(\beta) + \beta\nu\left(1 + \left[\ln\tfrac{l^*-l_0}{\beta\nu}\right]_+\right), \tag{2.8}$$

where $[a]_+ = \max\{a, 0\}$. Moreover,

$$l^* - l_0 \;\leq\; \left[\sqrt{l^*(\beta) - l_0} + \sqrt{\beta\nu}\right]^2. \tag{2.9}$$

Proof:
First part of inequality (2.8) follows from definitions (2.1) and (2.7). Let us prove the second part. Consider arbitrary $y^* \in \operatorname{Arg}\max\limits_{x \in P} l(x)$. Denote

$$y(\alpha) = x_0 + \alpha(y^* - x_0), \quad \alpha \in [0, 1].$$

In view of inequality (2.3.3) in [16], we have

$$F(y(\alpha)) \;\leq\; F(x_0) - \nu\ln(1 - \alpha), \quad \alpha \in [0, 1).$$

Since $l(\cdot)$ is linear, this relation implies that

$$\begin{aligned}
l^*(\beta) \;&\geq\; \max_{\alpha \in [0,1)}\{l(y(\alpha)) - \beta[F(y(\alpha)) - F(x_0)]\} \\[4pt]
&\geq\; (1 - \alpha)l_0 + \alpha l^* + \beta\nu\ln(1 - \alpha), \quad \alpha \in [0, 1).
\end{aligned} \tag{2.10}$$

Maximum of the right-hand side of this inequality in α is attained at $\alpha^* = \left[1 - \frac{\beta\nu}{l^*-l_0}\right]_+$. Thus, if $\frac{l^*-l_0}{\beta\nu} \leq 1$ (that is $\alpha^* = 0$), then $l^* \leq l_0 + \beta\nu$, and (2.8) follows from (2.7). If $\alpha^* > 0$, then we get (2.8) by direct substitution.

On the other hand, from (2.10) we have

$$l^* - l_0 \;\leq\; \tfrac{1}{\alpha}\left[l^*(\beta) - l_0 + \beta\nu\ln\left(1 + \tfrac{\alpha}{1-\alpha}\right)\right] \;\leq\; \tfrac{1}{\alpha}[l^*(\beta) - l_0] + \tfrac{\beta\nu}{1-\alpha}.$$

Minimizing the latter expression in α, we get (2.9). $\quad\square$

Corollary 1 *For any $\beta > 0$ we have*

$$l^* \;\leq\; l^*(\beta) + \beta\nu\left[1 + 2\ln\left(1 + \sqrt{\tfrac{l^*(\beta)-l_0}{\beta\nu}}\right)\right]. \tag{2.11}$$

3   Barrier subgradient scheme

In this paper we consider convex optimization problems in the following form

$$\text{Find } f_* \overset{\text{def}}{=} \max_x \{f(x): \ x \in P\}, \tag{3.1}$$

where f is a concave function and P satisfies the structural assumptions specified in the beginning of Section 2. In the sequel, we assume f to be subdifferentiable on P_0 and the set P to be simple. The latter means that the auxiliary optimization problem (2.2) can be easily solved.

Consider now the generic scheme of *Barrier Subgradient Method* (BSM).

$$\boxed{\begin{array}{l} \textbf{Initialization:} \text{ Set } s_0 = 0 \in E^*. \\[1.5em] \hline \\[-0.5em] \textbf{Iteration } (k \geq 0)\textbf{:} \\[1em] \textbf{1.} \text{ Choose } \beta_k > 0 \text{ and compute } x_k = u^*_{\beta_k}(s_k). \\[1em] \textbf{2.} \text{ Choose } \lambda_k > 0 \text{ and set } s_{k+1} = s_k + \lambda_k \nabla f(x_k). \end{array}} \tag{3.2}$$

In order to analyze performance of this method, consider the following *gap functions*:

$$l_k(y) \quad = \quad \sum_{i=0}^{k} \lambda_i \langle \nabla f(x_i), y - x_i \rangle,$$

$$l_k^* \quad \overset{\text{def}}{=} \quad \max_{y \in P} l_k(y), \quad k \geq 0.$$

Theorem 1 *Assume that the parameters of scheme (3.2) satisfy condition*

$$\lambda_k \|\nabla f(x_k)\|^*_{x_k} \ < \ \beta_k \ \leq \ \beta_{k+1}, \quad k \geq 0. \tag{3.3}$$

Denote $S_k = \sum_{i=0}^{k} \lambda_i$, *and* $A_k = \sum_{i=0}^{k} \beta_i \omega_* \left(\frac{\lambda_i}{\beta_i} \|\nabla f(x_i)\|^*_{x_i} \right)$. *Then, for any* $k \geq 0$ *we have*

$$l_k^* \ \leq \ A_k + \beta_{k+1} \nu \left[1 + 2 \ln \left(1 + \sqrt{\frac{A_k}{\beta_{k+1}\nu} + c(Q) \frac{S_k}{\beta_{k+1}} \|\nabla f(x_0)\|^*_{x_0}} \right) \right], \tag{3.4}$$

where $c(Q) = 1$ *if* Q *is a cone and* $F(\cdot)$ *is a logarithmically homogeneous barrier, and* $c(Q) = 3$ *otherwise.*

Proof:
Note that for any $k \geq 0$ we have

$$U_{\beta_{k+1}}(s_{k+1}) \overset{(3.3)}{\leq} U_{\beta_k}(s_{k+1})$$

$$\overset{(2.4)}{\leq} U_{\beta_k}(s_k) + \lambda_k \langle \nabla f(x_k), u^*_{\beta_k}(s_k) - x_0 \rangle + \beta_k \omega_* \left(\frac{\lambda_k}{\beta_k} \|\nabla f(x_k)\|^*_{x_k} \right).$$

Since $U_{\beta_0}(0) = 0$, we conclude that

$$\langle s_{k+1}, x_{k+1} - x_0 \rangle - \beta_{k+1}[F(x_{k+1}) - F(x_0)] = U_{\beta_{k+1}}(s_{k+1})$$

$$\leq \sum_{i=0}^{k} \lambda_i \langle \nabla f(x_i), x_i - x_0 \rangle + \sum_{i=0}^{k} \beta_i \omega_* \left(\frac{\lambda_i}{\beta_i} \|\nabla f(x_i)\|_{x_i}^* \right). \tag{3.5}$$

In view of the first-order optimality condition for (2.2), for all $y \in P_0$ we have

$$\langle s_{k+1}, y - x_{k+1} \rangle \leq \beta_{k+1} \langle \nabla F(x_{k+1}), y - x_{k+1} \rangle. \tag{3.6}$$

Hence, for any $y \in P_0$ we obtain

$$\sum_{i=0}^{k} \lambda_i \langle \nabla f(x_i), y - x_i \rangle \overset{(3.5)}{\leq} \langle s_{k+1}, y - x_{k+1} \rangle + \beta_{k+1}[F(x_{k+1}) - F(x_0)] + A_k$$

$$\overset{(3.6)}{\leq} \beta_{k+1}[F(x_{k+1}) + \langle \nabla F(x_{k+1}), y - x_{k+1} \rangle - F(x_0)] + A_k$$

$$\leq \beta_{k+1}[F(y) - F(x_0)] + A_k.$$

Hence, $l_k^*(\beta_{k+1}) \leq A_k$. On the other hand, since f is concave, we obtain

$$l_k(x_0) = \sum_{i=0}^{k} \lambda_i \langle \nabla f(x_i), x_0 - x_i \rangle \geq \sum_{i=0}^{k} \lambda_i \langle \nabla f(x_0), x_0 - x_i \rangle$$

$$\geq -\|\nabla f(x_0)\|_{x_0}^* \cdot \sum_{i=0}^{k} \lambda_i \|x_0 - x_i\|_{x_0}.$$

In view of definition (2.1), we have $\langle \nabla F(x_0), x_i - x_0 \rangle \geq 0$. Hence, by Theorem 4.2.5 [8], $\|x_i - x_0\|_{x_0} \leq \nu + 2\sqrt{\nu} \leq 3\nu$ (recall that $\nu \geq 1$, e.g. Lemma 4.3.1 [8]). If Q is a cone, then $\|x_i - x_0\|_{x_0} \leq \nu$ (see Lemma 5 [17]). Thus, in any case

$$\|x_i - x_0\|_{x_0} \leq \nu \cdot c(Q), \quad i = 0, \dots, k,$$

and we conclude that $l_k(x_0) \geq -\nu \cdot c(Q) S_k \|\nabla f(x_0)\|_{x_0}^*$. Using our observations and inequality (2.11), we obtain (3.4). $\qquad\square$

Let us estimate now the rate of convergence of method (3.2) as applied to a specific problem class.

Definition 1 *We say that $f \in \mathcal{B}_M(P)$ if $\|\nabla f(x)\|_x^* \leq M$ for any $x \in P_0$.*

For function $f \in \mathcal{B}_M(P)$, we suggest the following values of parameters in (3.2):

$$\lambda_k = 1, \ k \geq 0, \qquad \beta_0 = \beta_1, \quad \beta_k = M \cdot \left(1 + \sqrt{\frac{k}{\nu}} \right), \ k \geq 1. \tag{3.7}$$

Theorem 2 *Let problem (3.1) with $f \in \mathcal{B}_M(P)$ be solved by method (3.2) with parameters given by (3.7). Then for any $k \geq 0$ we have*

$$\frac{1}{S_k} l_k^* \leq 2M \cdot \left(\sqrt{\frac{\nu}{k+1}} + \frac{\nu}{k+1} \right) \cdot \left(1 + \ln \left(2 + \frac{1}{2} c(Q) \sqrt{\nu(k+1)} \right) \right). \tag{3.8}$$

Proof:

Define $\tau_k = \frac{1}{M}\beta_k > 1$. In view of the choice of parameters (3.7) and assumptions of the theorem, we have $S_k = k+1$, and

$$
\begin{aligned}
A_k &= \sum_{i=0}^{k} \beta_i \omega_* \left(\frac{\lambda_i}{\beta_i} \|\nabla f(x_i)\|_{x_i}^* \right) \leq M \sum_{i=0}^{k} \tau_i \omega_* \left(\frac{1}{\tau_i} \right) \leq \tfrac{1}{2} M \sum_{i=0}^{k} \tau_i \frac{\tau_i^{-2}}{1-\tau_i^{-1}} \\
&= \tfrac{1}{2} M \sum_{i=0}^{k} \frac{1}{\tau_i - 1} = \frac{\sqrt{\nu}}{2} M \left[1 + \sum_{i=1}^{k} \frac{1}{\sqrt{i}} \right] \leq \sqrt{\nu} M \left[\tfrac{1}{2} + \sqrt{k} \right].
\end{aligned}
\tag{3.9}
$$

Further, by the same reasons

$$
\frac{S_k}{\beta_{k+1}} \|\nabla f(x_0)\|_{x_0}^* \leq \frac{k+1}{1+\sqrt{\frac{k+1}{\nu}}} \leq \sqrt{\nu(k+1)},
$$

$$
\frac{A_k}{\beta_{k+1}\nu} \leq \frac{\tfrac{1}{2}+\sqrt{k}}{\sqrt{\nu}+\sqrt{k+1}} \leq 1.
$$

Thus, substituting the above estimates in inequality (3.4), we obtain

$$
\begin{aligned}
\frac{l_k^*}{S_k} &\leq M \left[\frac{\sqrt{\nu}}{k+1} \left(\tfrac{1}{2} + \sqrt{k} \right) + \frac{\nu + \sqrt{\nu(k+1)}}{k+1} \left(1 + 2\ln \left(1 + \sqrt{1 + c(Q)\sqrt{\nu(k+1)}} \right) \right) \right] \\
&\leq 2M \cdot \left(\sqrt{\frac{\nu}{k+1}} + \frac{\nu}{k+1} \right) \cdot \left(1 + \ln \left(2 + \tfrac{1}{2}c(Q)\sqrt{\nu(k+1)} \right) \right).
\end{aligned}
$$

$\square$

With parameters chosen by (3.7), the scheme of method (3.2) can be written in the following form:

$$
x_{k+1} = \arg\max_{x \in P_0} \left\{ \frac{1}{k+1} \sum_{i=0}^{k} \langle \nabla f(x_i), x - x_i \rangle - M \frac{\sqrt{\nu}+\sqrt{k+1}}{\sqrt{\nu(k+1)}} \left[F(x) - F(x_0) \right] \right\}, \quad k \geq 0.
\tag{3.10}
$$

Since f is a concave function,

$$
\begin{aligned}
\frac{1}{S_k} l_k^* &= \frac{1}{S_k} \max_{y \in P} \sum_{i=0}^{k} \lambda_i \langle \nabla f(x_i), y - x_i \rangle \\
&\geq \frac{1}{S_k} \max_{y \in P} \sum_{i=0}^{k} \lambda_i [f(y) - f(x_i)] = f_* - \frac{1}{S_k} \sum_{i=0}^{k} \lambda_i f(x_i).
\end{aligned}
$$

Thus, the estimate (3.8) justifies the following rate of convergence for *primal* variables:

$$
f_* - \frac{1}{S_k} \sum_{i=0}^{k} \lambda_i f(x_i) \leq 2M \cdot \left(\sqrt{\frac{\nu}{k+1}} + \frac{\nu}{k+1} \right) \cdot \left(1 + \ln \left(2 + \tfrac{1}{2}c(Q)\sqrt{\nu(k+1)} \right) \right). \tag{3.11}
$$

Note that the value l_k^* is computable. Hence, it can be used for terminating the process.

To conclude this section, let us show that method (3.10) can also generate approximate solutions to a dual problem. For that, we need to employ an *internal structure* of our problem. Let us assume that it can be represented in a *saddle-point form*:

$$
f(x) = \min_{w \in S} \Psi(w, x) \rightarrow \max_{x \in P}, \tag{3.12}
$$

8

where $S \subset E_1$ is a closed convex set, and function $\Psi(w, x)$ is convex in $w \in S$ and concave and subdifferentiable in $x \in P$. Then, the dual problem is defined as

$$\text{Find } f_* = \min_{w \in S} \eta(w),$$

$$\eta(w) = \max_{y \in P} \Psi(w, y). \tag{3.13}$$

Since P is bounded, the above problem is well defined. Without loss of generality, it is always possible to choose

$$\nabla f(x) = \nabla_2 \Psi(w(x), x)$$

with some $w(x) \in \text{Arg} \min_{w \in S} \Psi(w, x) \subseteq S$.

Let us assume that $w(x)$ is computable for any $x \in P$. For the sake of completeness, we provide the following standard result with a simple proof.

Lemma 3 *Define $\bar{w}_k = \frac{1}{S_k} \sum\limits_{i=0}^{k} \lambda_i w(x_i)$, and $\bar{x}_k = \frac{1}{S_k} \sum\limits_{i=0}^{k} \lambda_i x_i$. Then*

$$\eta(\bar{w}_k) - f(\bar{x}_k) \leq \frac{1}{S_k} l_k^*. \tag{3.14}$$

Proof:
Since Ψ is concave in the second argument, for any $y \in P$ we have

$$\langle \nabla f(x_i), y - x_i \rangle = \langle \nabla_2 \Psi(w(x_i), x_i), y - x_i \rangle$$

$$\geq \Psi(w(x_i), y) - \Psi(w(x_i), x_i) = \Psi(w(x_i), y) - f(x_i).$$

Hence,

$$\frac{1}{S_k} l_k^* = \frac{1}{S_k} \max_{y \in P} \sum_{i=0}^{k} \lambda_i \langle \nabla f(x_i), y - x_i \rangle \geq \frac{1}{S_k} \max_{y \in P} \sum_{i=0}^{k} \lambda_i [\Psi(w(x_i), y) - f(x_i)]$$

$$\geq \max_{y \in P} \Psi(\bar{w}_k, y) - \frac{1}{S_k} \sum_{i=0}^{k} \lambda_i f(x_i) = \eta(\bar{w}_k) - \frac{1}{S_k} \sum_{i=0}^{k} \lambda_i f(x_i) \geq \eta(\bar{w}_k) - f(\bar{x}_k).$$

$\square$

Thus, the scheme (3.10) can indeed generate approximate primal-dual solutions:

$$\eta(\bar{w}_k) - f(\bar{x}_k) \leq 2M \cdot \left(\sqrt{\frac{\nu}{k+1}} + \frac{\nu}{k+1} \right) \cdot \left(1 + \ln \left(2 + \frac{1}{2} c(Q) \sqrt{\nu(k+1)} \right) \right). \tag{3.15}$$

4 Maximizing positive concave functions

Consider now a convex optimization problem

$$\text{Find } \psi_* \stackrel{\text{def}}{=} \max_{x} \{ \psi(x) : x \in P \}, \tag{4.1}$$

where the set $P = \hat{P} \bigcap Q$ satisfies assumptions introduced for problem (3.1). However, now we assume that function ψ is concave and *non-negative* on Q:

$$\psi(x) > 0, \quad \forall x \in \text{int} \, Q. \tag{4.2}$$

Lemma 4 *Let ψ be concave and non-negative on Q. Then for any $x \in \text{int}\, Q$ we have*

$$\|\nabla\psi(x)\|_x^* \leq \psi(x). \tag{4.3}$$

Proof:
Let us choose an arbitrary $x \in \text{int}\, Q$ and $r \in [0,1)$. Define

$$y = x - \tfrac{r}{\|\nabla\psi(x)\|_x^*}[\nabla^2 F(x)]^{-1}\nabla\psi(x).$$

In view of Theorem 4.1.5 [8], $y \in \text{int}\, Q$. Therefore,

$$0 \leq \psi(y) \leq \psi(x) + \langle \nabla\psi(x), y - x \rangle = \psi(x) - r\|\nabla\psi(x)\|_x^*.$$

Since r is an arbitrary value from $[0,1)$, we get (4.3). $\qquad\square$

This result has an important corollary. Let us apply to the objective function of problem (4.1) a logarithmic transformation:

$$f(x) \overset{\text{def}}{=} \ln\psi(x). \tag{4.4}$$

Lemma 5 *Let ψ be concave and positive in the sense of (4.2). Then $f \in \mathcal{B}_1(Q)$, and it is concave on Q.*

Proof:
Indeed, it is well known that the logarithm of concave function is concave. It remains to note that $\nabla f(x) = \frac{1}{\psi(x)}\nabla\psi(x)$ and apply inequality (4.3). $\qquad\square$

Thus, in order to solve problem (4.1), we can apply method (3.2) to problem (3.1) with objective function defined by (4.4). The resulting optimization scheme looks as follows:

$$x_{k+1} = \arg\max_{x \in P_0}\left\{ \tfrac{1}{k+1}\sum_{i=0}^{k}\langle \tfrac{\nabla\psi(x_i)}{\psi(x_i)}, x - x_i \rangle - \tfrac{\sqrt{\nu}+\sqrt{k+1}}{\sqrt{\nu(k+1)}}[F(x) - F(x_0)]\right\}, \quad k \geq 0. \tag{4.5}$$

For scheme (4.5), we can guarantee a certain rate of convergence in *relative scale*.

Theorem 3 *Let sequence $\{x_k\}_{k=0}^{\infty}$ be generated by method (4.5) for problem (4.1). Then for any $k \geq 0$ we have*

$$\left[\prod_{i=0}^{k}\psi(x_i)\right]^{\frac{1}{k+1}} \geq \psi_* \cdot \exp\left\{-2\left(\sqrt{\tfrac{\nu}{k+1}} + \tfrac{\nu}{k+1}\right)\left(1 + \ln\left(2 + \tfrac{c(Q)}{2}\sqrt{\nu(k+1)}\right)\right)\right\}$$

$$\geq \psi_* \cdot \left[1 - 2\left(\sqrt{\tfrac{\nu}{k+1}} + \tfrac{\nu}{k+1}\right)\left(1 + \ln\left(2 + \tfrac{c(Q)}{2}\sqrt{\nu(k+1)}\right)\right)\right]. \tag{4.6}$$

Proof:
Indeed, we just apply method (3.10) to function f defined by (4.4). Since $f \in \mathcal{B}_1(Q) \subseteq \mathcal{B}_1(P)$, by (3.8) we conclude that

$$f_* - \tfrac{1}{k+1}\sum_{i=0}^{k}f(x_i) \leq \delta_k \overset{\text{def}}{=} 2\left(\sqrt{\tfrac{\nu}{k+1}} + \tfrac{\nu}{k+1}\right)\left(1 + \ln\left(2 + \tfrac{c(Q)}{2}\sqrt{\nu(k+1)}\right)\right).$$

Hence, $\left[\prod_{i=0}^{k}\psi(x_i)\right]^{\frac{1}{k+1}} \geq \psi_* \cdot e^{-\delta_k} \geq \psi_* \cdot (1-\delta_k)$. This is exactly (4.6). $\qquad\square$

Let us show how we can treat a problem dual to (4.1). For simplicity, assume that

$$\psi(x) = \min_{u\in\Omega} \Psi_0(u,x), \tag{4.7}$$

where $\Omega \subset E_1$ is a closed convex set. In this case, condition (4.2) can be written as

$$\Psi_0(u,x) \geq 0, \quad u \in \Omega, \; x \in P. \tag{4.8}$$

Note that

$$\max_{x\in P} \ln \psi(x) = \max_{x\in P} \min_{\tau>0} \min_{u\in\Omega} \left[\tau\Psi_0(u,x) - \ln\tau - 1\right]$$

$$= \max_{x\in P} \min_{\substack{v\in\tau\Omega,\\ \tau>0}} \left[\tau\Psi_0\left(\tfrac{1}{\tau}v,x\right) - \ln\tau - 1\right]$$

$$= \min_{\substack{v\in\tau\Omega,\\ \tau>0}} \left\{\eta(w) \equiv \eta(v,\tau) \stackrel{\text{def}}{=} -1 - \ln\tau + \tau\psi^*\left(\tfrac{1}{\tau}v\right)\right\},$$

where $\psi^*(u) = \max_{x\in P} \Psi_0(u,x)$.

Denote by $u(x)$ a solution of the minimization problem (4.7). Then $w(x)$ is clearly defined as follows

$$w(x) = (v(x), \tau(x)), \quad v(x) = \tau(x)u(x), \quad \tau(x) = \tfrac{1}{\psi(x)}.$$

In accordance to Lemma 3, we can form $\bar{w}_k = (\bar{v}_k, \bar{\tau}_k)$ with

$$\bar{v}_k = \tfrac{1}{k+1}\sum_{i=0}^{k}\tfrac{u(x_i)}{\psi(x_i)}, \quad \bar{\tau}_k = \tfrac{1}{k+1}\sum_{i=0}^{k}\tfrac{1}{\psi(x_i)}.$$

Denote $\bar{x}_k = \tfrac{1}{k+1}\sum_{i=0}^{k}x_i$, and $\bar{u}_k = \tfrac{\bar{v}_k}{\bar{\tau}_k} = \sum_{i=0}^{k}\tfrac{u(x_i)}{\psi(x_i)} / \left[\sum_{i=0}^{k}\tfrac{1}{\psi(x_i)}\right] \in \Omega$. Then, by (3.14) we get

$$\tfrac{1}{S_k}l_k^* \geq \eta(\bar{w}_k) - \ln\psi(\bar{x}_k) = -1 - \ln\bar{\tau}_k + \bar{\tau}_k\psi^*\left(\tfrac{1}{\bar{\tau}_k}\bar{v}_k\right) - \ln\psi(\bar{x}_k)$$

$$= -1 - \ln\bar{\tau}_k + \bar{\tau}_k\psi^*\left(\bar{u}_k\right) - \ln\psi(\bar{x}_k) \geq \ln\tfrac{\psi^*(\bar{u}_k)}{\psi(\bar{x}_k)}.$$

Hence,

$$\psi(\bar{x}_k) \geq \psi^*(\bar{u}_k) \cdot \exp\left\{-\tfrac{1}{S_k}l_k^*\right\}. \tag{4.9}$$

5 Applications

In this section we are going to consider several application examples for method (4.5). We will say that some value $\bar{\phi}$ is a δ-approximation of the optimal value $\phi_* > 0$ in *relative scale* if

$$\phi_* \geq \bar{\phi} \geq \phi_* \cdot e^{-\delta}, \quad \delta > 0.$$

In the complexity estimates, a short notation $O^*(\cdot)$ is used for indicating that the logarithmic factors are omitted.

5.1　Fractional covering problem

Consider the following *fractional covering* problem:

$$\text{Find } \phi_* \overset{\text{def}}{=} \min_y \{ \langle b, y \rangle : \ A^T y \ge c, \ y \ge 0 \in R^m \}, \tag{5.1}$$

where $A = (a_1, \ldots, a_n)$ is an $(m \times n)$-matrix with non-negative coefficients, and vectors $b \in R^m$ and $c \in R^n$ have positive coefficients. Denote

$$\psi(y) \ = \ \min_{1 \le i \le n} \tfrac{1}{c^{(i)}} \langle a_i, y \rangle.$$

Note that ψ is concave and homogeneous of degree one. Therefore

$$\phi_* \ = \ \min_y \left\{ \tfrac{\langle b, y \rangle}{\psi(y)} : \ y \ge 0 \in R^m \right\}$$

$$= \ \left[\max_y \left\{ \tfrac{\psi(y)}{\langle b, y \rangle} : \ y \ge 0 \in R^m \right\} \right]^{-1}$$

$$= \ \left[\max_y \left\{ \psi(y) : \ \langle b, y \rangle = 1, \ y \ge 0 \in R^m \right\} \right]^{-1}.$$

Thus, problem (5.1) can be written in the form (4.1) with $K = R_+^m$,

$$F(y) \ = \ - \sum_{j=1}^m \ln y^{(j)}, \quad \nu = m,$$

and $\hat{P} = \{y : \ \langle b, y \rangle = 1\}$. Hence, in accordance to the estimate (4.6) a δ-approximation of $\phi_* = \psi_*^{-1}$ in relative scale can be found in $O^*(\tfrac{m}{\delta^2})$ iterations of the method (4.5). Each iteration of the scheme needs $O(mn)$ operations for computing $\psi(y)$ and its subgradient, and essentially $O(m \ln m)$ operation for solving the auxiliary maximization problem in (4.5) (see Appendix). Of course, this computational strategy is reasonable if $m << n$. Otherwise, it is better to solve the dual form of problem (5.1) by smoothing technique [15].

5.2　Maximal concurrent flow problem

Consider a network consisting of the set of nodes $\mathcal{N}$, $|\mathcal{N}| = n$, and the set of directed arcs

$$\mathcal{A} \ = \ \{ \alpha = (i,j), \ i, j \in \mathcal{N} \}, \quad |\mathcal{A}| = m.$$

We assume that all arcs have bounded capacities. Formally this means that the arc flow vector $f \in R_+^m$ must satisfy the capacity constraint:

$$f \le \bar{f}.$$

Let us introduce the set of origin-destination pairs

$$\mathcal{OD} \ = \ \{ (i,j), \ i, j \in \mathcal{N} \}.$$

Each pair $(i,j) \in \mathcal{OD}$ generates for nodes i and j a directed flow $f_{i,j} \in R_+^m$ of level $d_{i,j}$. Formally, this means that the vectors $f_{i,j}$ must satisfy the system of linear equatuions

$$B f_{i,j} = d_{i,j}(e_i - e_j), \quad (i,j) \in \mathcal{OD},$$

where B is the balance matrix of the network and $e_{(\cdot)}$ is the corresponding unit vectors in R^n.

The *maximal concurrent flow* problem can be written as follows

$$\text{Find } \lambda_* \stackrel{\text{def}}{=} \max_{\lambda f, f_{i,j}} \{\lambda : \; B f_{i,j} = \lambda \cdot d_{i,j}(e_i - e_j), \quad f_{i,j} \geq 0, \quad (i,j) \in \mathcal{OD},$$
$$\sum_{(i,j) \in \mathcal{OD}} f_{i,j} \leq \bar{f} \, \}. \tag{5.2}$$

It is well known [19] that this problem can be written in a dual form:

$$\psi_* \stackrel{\text{def}}{=} \lambda_*^{-1} = \max_t \{\psi(t) : \; \langle \bar{f}, t \rangle = 1, \; t \geq 0 \in R^m\},$$

$$\psi(t) = \sum_{(i,j) \in \mathcal{OD}} d_{i,j} \cdot SP_{i,j}(t), \tag{5.3}$$

where the function $SP_{i,j}(t)$ is the *shortest path distance* between nodes i and j with respect to a non-negative arc travel time vector $t \in R^m$.

Clearly, function ψ in (5.3) satisfies all assumptions introduced for problem (4.1). Therefore (5.3) can be treated by method (4.5). In accordance to the estimate (4.6), a δ-approximation of ψ_* in relative scale can be found in $O^*(\frac{m}{\delta^2})$ iterations. Each iteration of the scheme needs a computation of shortest-path distances for all origin-destination pairs. Complexity of solving the auxiliary maximization problem in (4.5) is essentially $O(m \ln m)$ operations (see Appendix). Note that we are also able to reconstruct the dual solutions (origin-destination flows) using the technique described in the end of Section 4.

5.3 Minimax problem with nonnegative components

Consider the following minimax problem:

$$\text{Find } \psi_* \stackrel{\text{def}}{=} \min_x \left\{ \max_{1 \leq i \leq m} f_i(x) : \; x \in S \right\}, \tag{5.4}$$

where S is a closed convex set and all functions $f_i(x)$ are convex and non-negative on S. We assume that the function

$$\psi(y) = \min_{x \in S} \left\{ \sum_{i=1}^m y^{(i)} f_i(x) \right\}$$

is well defined for any $y \geq 0 \in R^m$. Moreover, we assume that the values of this function and its subgradients are easily computable.

Then we can rewrite problem (5.4) in the dual form

$$\psi_* = \max_y \{\psi(y) : \; \langle e, y \rangle = 1, \; y \geq 0 \in R^m\}, \tag{5.5}$$

where $e \in R^m$ is the vector of all ones.

Note that (5.5) satisfies all assumption of problem (4.1). Therefore, in accordance to the estimate (4.6), an δ-approximation of ψ_* in relative scale can be found by method (4.5) in $O^*(\frac{m}{\delta^2})$ iterations. Each iteration of the scheme results in minimization of the weighted sum of functions f_i.

13

5.4 Semidefinite relaxation of boolean quadratic problem

Consider the following maximization problem:

$$\text{Find } f_* \stackrel{\text{def}}{=} \max_x \{ \langle Ax, x \rangle : x^{(i)} = \pm 1, \ i = 1, \dots, n \}, \tag{5.6}$$

where A is a symmetric positive definite $(n \times n)$-matrix. Clearly, this problem is NP-hard. However, it is well known that its optimal value can be approximated in polynomial time with certain dimension-independent relative accuracy (see [7]). Namely, define

$$\psi_* = \min_y \{ \langle e, y \rangle : D(y) \succeq A \}, \tag{5.7}$$

where $D(y)$ is a diagonal $(n \times n)$-matrix with vector y on the diagonal. Then

$$\tfrac{2}{\pi} \psi_* \le f_* \le \psi_*.$$

Usually the problem (5.7) is treated by Interior-Point Methods. However, note that quite often it is useless to compute an approximation to ψ_* with a high relative accuracy. Therefore it seems reasonable to solve it by a cheap gradient scheme.

Let us justify another representation for ψ_*.

Lemma 6 *Let $A = L^T L$. Then*

$$\psi_* = \max_X \left\{ \psi(X) \stackrel{\text{def}}{=} \left[\sum_{i=1}^n \langle X q_i, q_i \rangle^{1/2} \right]^2 : \langle I, X \rangle = 1, \ X \succeq 0 \right\}, \tag{5.8}$$

where q_i are the columns of the matrix L, I is the identity matrix, and the scalar product in the space of symmetric matrices is defined in a natural way.

Proof:
Indeed, since $A \succ 0$, we have

$$\psi_* = \min_u \left\{ \sum_{i=1}^n \frac{1}{u^{(i)}} : A^{-1} \succeq D(u) \right\}$$

$$= \min_u \max_{Y \succeq 0} \left\{ \sum_{i=1}^n \frac{1}{u^{(i)}} + \langle Y, D(u) - A^{-1} \rangle \right\}$$

$$= \max_{Y \succeq 0} \min_u \left\{ \sum_{i=1}^n \left(\frac{1}{u^{(i)}} + Y^{(i,i)} u^{(i)} \right) - \langle Y, A^{-1} \rangle \right\}$$

Thus, $\psi_* = \max_{Y \succeq 0} \left\{ 2 \sum_{i=1}^n \left[Y^{(i,i)} \right]^{1/2} - \langle Y, A^{-1} \rangle \right\}$. Maximizing the objective function in this problem along a fixed direction $Y \ge 0$, we obtain

$$\psi_* = \max_{Y \succeq 0} \left\{ \frac{1}{\langle Y, A^{-1} \rangle} \left[\sum_{i=1}^n \left[Y^{(i,i)} \right]^{1/2} \right]^2 \right\}.$$

Choosing in this problem new variables $X = L^{-T} Y L^{-1}$, we obtain representation (5.8). $\square$

14

Note that function ψ in (5.8) is concave. Moreover, it is differentiable and positive at any $X \succ 0$. In our case, Q is the cone of positive-semidefinite matrices with

$$F(X) \;=\; -\ln \det X, \quad \nu = n.$$

Hence, (5.8) satisfies conditions of the problem (4.1). Consequently, ψ_* can be approximated by (4.5) in $O^*(\frac{n}{\delta^2})$ iterations, where δ is the desired relative accuracy. In our case, each iteration of the scheme (4.5) requires a Cholesky decomposition of $(n \times n)$-matrix.

6 Online optimization as an alternative to Stochastic Programming

6.1 Decision-making in uncertain environment

Consider a repeatable decision-making process with uncertain income. Assume we have $N + 1$ periods of time, each of which corresponds to a *full* production cycle. In the beginning of kth period we choose a production strategy

$$x_k \in P, \quad k = 0, \dots, N,$$

where the structure of P satisfies the assumptions of Section 2. The results of different economic activities in this period are given by *production function* $\psi_k(x) \geq 0$, $x \in P$. The value $\psi_k(x)$ is equal to the *rate of growth* of the capital invested at the beginning of period k in accordance to production strategy $x \in P$. Function $\psi_k(\cdot)$ becomes known only in the end of the period k. So, it can be used for choosing the production strategies of the next periods.

Assume for a moment that we know in advance all production functions

$$\psi_k(x), \quad k = 0, \dots, N.$$

However, by certain conditions we are obliged to apply in all these periods the same strategy $x \in P$. In this case, of course, it is reasonable to use

$$x_N^* \;\overset{\text{def}}{=}\; \arg\max_{x \in P} \prod_{k=0}^{N} \psi_k(x).$$

Then, the average efficiency of this static strategy is given by

$$\psi_N^* \;=\; \left[\prod_{k=0}^{N} \psi_k(x^*) \right]^{\frac{1}{N+1}}.$$

However, usually the future is unknown. Instead, we have a freedom in choosing for each period k a specific production strategy $x_k \in P$. Let us look at its possible efficiency.

Suppose we know a ν-self-concordant barrier $F(x)$ for the set Q. Then, we could apply the following variant of method (4.5):

$$x_{k+1} = \arg\max_{x \in P_0} \left\{ \frac{1}{k+1} \sum_{i=0}^{k} \langle \frac{\nabla \psi_i(x_i)}{\psi_i(x_i)}, x - x_i \rangle - \frac{\sqrt{\nu}+\sqrt{k+1}}{\sqrt{\nu}(k+1)} \left[F(x) - F(x_0) \right] \right\}, \quad k \geq 0. \quad (6.1)$$

In this case, after $N + 1$ periods, the average rate of growth is given by

$$\Psi_N \stackrel{\text{def}}{=} \left[\prod_{k=0}^{N} \psi_k(x_k) \right]^{\frac{1}{N+1}}.$$

Theorem 4 *For any $N \geq 0$ we have $\Psi_N \geq \psi_N^* \cdot e^{-\delta_N}$ with*

$$\delta_N = 2 \left(\sqrt{\tfrac{\nu}{N+1}} + \tfrac{\nu}{N+1} \right) \cdot \left(1 + \ln \left(2 + \tfrac{c(Q)}{2} \sqrt{\nu(N+1)} \right) \right) \rightarrow 0$$

as $N \to \infty$.

Proof:
The proof is very similar to proofs of Theorems 1 and 2. Denote

$$f_k(x) = \ln \psi_k(x), \quad f(x) = \tfrac{1}{N+1} \sum_{k=0}^{N} f_k(x), \quad s_k = \sum_{i=0}^{k} \nabla f_i(x_i) = \sum_{i=0}^{k} \tfrac{\nabla \psi_i(x_i)}{\psi_i(x_i)}.$$

Note that method (6.1) can be seen as an application of scheme (3.2), (3.7) to a varying objective function.

For any $k \geq 0$ we have

$$U_{\beta_{k+1}}(s_{k+1}) \leq U_{\beta_k}(s_{k+1})$$

$$\stackrel{(2.4)}{\leq} U_{\beta_k}(s_k) + \langle \nabla f_k(x_k), u_{\beta_k}^*(s_k) - x_0 \rangle + \beta_k \omega_* \left(\tfrac{1}{\beta_k} \| \nabla f_k(x_k) \|_{x_k}^* \right)$$

$$\stackrel{(4.3)}{\leq} U_{\beta_k}(s_k) + \langle \nabla f_k(x_k), u_{\beta_k}^*(s_k) - x_0 \rangle + \beta_k \omega_* \left(\tfrac{1}{\beta_k} \right).$$

Since $U_{\beta_0}(0) = 0$, we conclude that

$$\langle s_{N+1}, x_{N+1} - x_0 \rangle - \beta_{N+1}[F(x_{N+1}) - F(x_0)]$$

$$= U_{\beta_{N+1}}(s_{N+1}) \leq \sum_{i=0}^{N} \langle \nabla f_i(x_i), x_i - x_0 \rangle + \sum_{i=0}^{N} \beta_i \omega_* \left(\tfrac{1}{\beta_i} \right) \qquad (6.2)$$

$$\stackrel{(3.9)}{\leq} \sum_{i=0}^{N} \langle \nabla f_i(x_i), x_i - x_0 \rangle + \sqrt{\nu} \left[\tfrac{1}{2} + \sqrt{N} \right].$$

In view of the first-order optimality condition for (2.2), for all $y \in P_0$ we have

$$\langle s_{N+1}, y - x_{N+1} \rangle \leq \beta_{N+1} \langle \nabla F(x_{N+1}), y - x_{N+1} \rangle. \qquad (6.3)$$

Hence, using concavity of all functions f_i, for any $y \in P$ we obtain

$$l_N(y) \stackrel{\text{def}}{=} \sum_{i=0}^{N} \langle \nabla f_i(x_i), y - x_i \rangle$$

$$\stackrel{(6.2)}{\leq} \langle s_{N+1}, y - x_{N+1} \rangle + \beta_{N+1}[F(x_{N+1}) - F(x_0)] + \sqrt{\nu} \left[\tfrac{1}{2} + \sqrt{N} \right]$$

$$\stackrel{(6.3)}{\leq} \beta_{N+1}[F(x_{N+1}) + \langle \nabla F(x_{N+1}), y - x_{N+1} \rangle - F(x_0)] + \sqrt{\nu} \left[\tfrac{1}{2} + \sqrt{N} \right]$$

$$\leq \beta_{N+1}[F(y) - F(x_0)] + \sqrt{\nu} \left[\tfrac{1}{2} + \sqrt{N} \right].$$

Hence, $l_N^*(\beta_{N+1}) \leq \sqrt{\nu}\left[\frac{1}{2} + \sqrt{N}\right]$. On the other hand,

$$l_N(x_0) = \sum_{i=0}^{N}\langle \nabla f_i(x_i), x_0 - x_i\rangle \geq \sum_{i=0}^{N}\langle \nabla f_i(x_i), x_0 - x_i\rangle$$

$$\geq -\nu \cdot c(Q) \cdot (N+1).$$

Hence, by (3.4) $l_N^* \leq (N+1) \cdot \delta_N$. On the other hand,

$$\frac{1}{N+1} l_N^* = \frac{1}{N+1}\max_{y\in P}\left\{\sum_{i=0}^{N}\langle \nabla f_i(x_i), y - x_i\rangle\right\}$$

$$\geq \frac{1}{N+1}\max_{y\in P}\left\{\sum_{i=0}^{N}[f_i(y) - f_i(x_i)]\right\}$$

$$= \ln \psi_N^* - \ln \Psi_N.$$

$\square$

Let us look now at several applications of above theorem.

6.2 Portfolio management

Let $x \in \Delta_n$ be the structure of our portfolio. Denote by $c_k^{(i)} \geq 0$, $i = 1, \ldots, n$, the growth of the price of the stock i during the day $k \geq 0$. Then the optimal *constant rebalanced portfolio* is defined as

$$x_N^* = \arg\max_{x\in P}\prod_{k=0}^{N}\langle c_k, x\rangle, \quad \psi_N^* = \left[\prod_{k=0}^{N}\langle c_k, x_N^*\rangle\right]^{1/(N+1)}.$$

For the set $Q = R_+^n$, we can apply the standard n-self-concordant barrier

$$F(x) = -\sum_{i=1}^{n}\ln x^{(i)}.$$

Then, we can apply the following variant of method (6.1):

$$x_{k+1} = \arg\max_{x\in P_0}\left\{\frac{1}{k+1}\sum_{i=0}^{k}\frac{\langle c_i, x - x_i\rangle}{\langle c_i, x_i\rangle} - \frac{\sqrt{\nu} + \sqrt{k+1}}{\sqrt{\nu}(k+1)}[F(x) - F(x_0)]\right\}, \quad k \geq 0. \quad (6.4)$$

In this case, after $N + 1$ periods, the average rate of growth of our portfolio is given by

$$\Psi_N \overset{\text{def}}{=} \left[\prod_{k=0}^{N}\langle c_k, x_k\rangle\right]^{\frac{1}{N+1}}.$$

In view of Theorem 4, we have $\Psi_N \geq \psi_N^* \cdot e^{-\delta_N(n)}$. Note that each step of the algorithm (6.4) is implementable in $O(n\ln n)$ arithmetic operations (see Appendix).

6.3 Processes with full production cycles

Assume that in our economy there are n elastic production processes. In the beginning of kth period, we know the cost $a_k^{(i)} > 0$ for producing one unit of product i, $i = 1, \ldots, n$. This cost is derived from the prices of row materials, labor, equipment, etc. However, the price $b_k^{(i)} \geq 0$ of the unit of product i becomes known only in the end of period k, when we sell it. It may depend on competition in the market, uncertain preferences of the consumers, etc. Denoting by $x^{(i)}$ the fraction of the capital invested in the process i, we come to the following model:

$$
\begin{aligned}
\psi_k(x) &= \sum_{i=1}^{n} \frac{b_k^{(i)}}{a_k^{(i)}} \cdot x^{(i)}, \\
x &= (x^{(1)}, \ldots, x^{(n)})^T \in Q \stackrel{\text{def}}{=} R_+^n, \\
\hat{P} &= \Delta_n.
\end{aligned}
\tag{6.5}
$$

Then we can apply method (6.1) with

$$
F(x) = -\sum_{i=1}^{n} \ln x^{(i)}, \quad \nu = n.
$$

In this situation, the complexity of solving the auxiliary maximization problem in (6.1) is again $O(n \ln n)$ arithmetic operations (see Appendix).

6.4 Discussion

Without entering into a long discussion of the results of this section, let us mention that Theorem 4, being applied in an uncertain environment, delivers an *absolute and risk-free guarantee* for a certain level of efficiency of online optimization strategy (6.1). For obtaining such a result, we do not need to introduce the standard machinery related to random events, risk measures, stochastic or robust optimization. Since we compare the efficiency of *dynamic* strategy with a *static* one, this guarantee may look too weak. However, note that the solution of the standard one-stage stochastic programming problem

$$
x_* = \arg\max_{x \in P} \mathcal{E}_\zeta[f(x, \zeta)],
\tag{6.6}
$$

where $\mathcal{E}_\zeta[\cdot]$ denotes the expectation with respect to random vector ζ, is *static* by its origin (otherwise, maximization of *expectation* has no sense). At the same time, the model $f(x, \xi)$, constructed by the analysis of the *past*, can hardly overpass by its quality the advantages of knowing exactly the *future*. Of course, one could suggest to apply multi-stage stochastic models. However, from the best of our knowledge, these models are quite far from being numerically tractable.

From our point of view, the main drawback of online optimization strategy (6.1) is its low rate of convergence. Therefore, it is efficient only for the processes where the average gain is big as compared to the number of iterations and the parameter of the barrier function. An interesting application of this technique can be rather found in a long-run production planning and management, than in the stock market.

References

[1] A. d'Aspremont, O. Banerjee, L. El Ghaoui. First-Order Methods for Sparse Covariance Selection. *SIAM Journal on Matrix Analysis and its Applications*, **30**(1), 56-66 (2008).

[2] D. Bienstock. Potential Function Methods for Approximately Solving Linear Programming Problems: Theory and Practice. Springer, 2002.

[3] M.D. Grigoriadis, L.G. Khachiyan. Fast approximation schemes for convex programs with many blocks and coupling constraints. *SIAM J.Optim.* **4**, 86–107 (1994).

[4] M.D. Grigoriadis, L.G. Khachiyan. Approximate minimum-cost multicommodity flows. *Mathematical Programming*, **75**, 477–482 (1996).

[5] A. Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J.Optim.* **15**(1), 229–251 (2004).

[6] A. Nemirovsky and D. Yudin. Problem Complexity and Method Efficiency in Optimization. Wiley, 1983.

[7] Yu.Nesterov. Semidefinite Relaxation and Nonconvex Quadratic Optimization. *Optimization Methods and Software*, **9**, 141–160 (1998).

[8] Yu. Nesterov. *Introductory Lectures on Convex Optimization.* Kluwer, Boston, 2004.

[9] Yu. Nesterov. "Smooth minimization of non-smooth functions", *Mathematical Programming*, **103** (1), 127-152 (2005).

[10] Yu. Nesterov. Excessive gap technique in nonsmooth convex minimization. *SIAM J. Optim.* **16** (1), 235-249 (2005).

[11] Yu. Nesterov. Dual extrapolation and its application for solving variational inequalities and related problems. *Mathematical Programming*, **109**(2-3), 319-344 (2007).

[12] Yu. Nesterov. Smoothing technique and its applications in semidefinite optimization. *Mathematical Programming*, **110**(2), 245-259 (2007).

[13] Yu. Nesterov. Primal-dual subgradient methods for convex problems. *Mathematical Programming* DOI 10.1007/s10107-007-0149-x, (2007).

[14] Yu. Nesterov. Gradient methods for minimizing composite objective function. CORE Discussion Paper 2007/76, (2007).

[15] Yu. Nesterov. Rounding of convex sets and efficient gradient methods for linear programming problems. *Optimization Methods and Software*, **23**(1), 109-128 (2008).

[16] Yu. Nesterov, A. Nemirovskii. *Interior point polynomial methods in convex programming: Theory and Applications,*, SIAM, Philadelphia, 1994.

[17] Yu.Nesterov, M.J.Todd and Y.Ye. Infeasible-start Primal-Dual Methods and Infeasibility Detectors. *Mathematical Programming*, **84**(2), 227-267 (1999).

[18] S.A. Plotkin, D. Shmoys, and E. Tardos. Fast approximation algorithms for fractional packing and covering problems. *Mathematics of Operations Research*, **20**, 257–301 (1995).

[19] F. Shahrokhi, D. W. Matula, The maximum concurrent flow problem, Journal of the ACM, 37: 318-334, 1991.

7 Appendix: Complexity of barrier projection onto the standard simplex

In the case $K = R_+^n$, we can take

$$F(x) = -\sum_{i=1}^{n} \ln x^{(i)}, \quad \nu = n.$$

Consider $\hat{P} = \{x \in R^n : \langle e, x \rangle = 1\}$. Then, at each iteration of method (4.5) we need to solve the following problem:

$$\phi^* \stackrel{\text{def}}{=} \max_x \left\{ \langle s, x \rangle + \sum_{i=1}^{n} \ln x^{(i)} : \sum_{i=1}^{n} x^{(i)} = 1 \right\}. \tag{7.1}$$

Let us show that its complexity does not depend on the size of particular data (that is the coefficients of vector $s \in R^n$).

Consider the following Lagrangian:

$$\mathcal{L}(x, \lambda) = \langle s, x \rangle + \sum_{i=1}^{n} \ln x^{(i)} + \lambda \cdot \left[1 - \sum_{i=1}^{n} x^{(i)} \right], \quad x \in R^n, \ \lambda \in R.$$

The dual function

$$\phi(\lambda) = \max_x \left\{ \mathcal{L}(x, \lambda) : \sum_{i=1}^{n} x^{(i)} = 1 \right\} \stackrel{\text{def}}{=} \mathcal{L}(x(\lambda), \lambda)$$

is defined by the vector $x(\lambda) : x^{(i)}(\lambda) = \frac{1}{\lambda - s^{(i)}}$, $i = 1, \ldots, n$. Thus,

$$
\begin{aligned}
\phi(\lambda) &= -n + \lambda - \sum_{i=1}^{n} \ln \left(\lambda - s^{(i)} \right), \\
\phi_* &= \min_\lambda \left\{ \phi(\lambda) : \lambda > \max_{1 \le i \le n} s^{(i)} \right\}.
\end{aligned} \tag{7.2}
$$

Note that $\phi(\cdot)$ is a self-concordant function. Therefore we can apply to its minimization a standard Newton scheme, which converges quadratically starting from any λ from the region

$$\mathcal{Q}(s) = \{\lambda : 4(\phi'(\lambda))^2 \le \phi''(\lambda)\}.$$

(see, for example, Section 4.1.5 in [8]). Let us show that the complexity of finding a starting point from this set does not depend on initial data.

Consider the function $\psi(\lambda) = -\phi'(\lambda) = \sum_{i=1}^{n} \frac{1}{\lambda - s^{(i)}} - 1$. Clearly, the problem (7.2) is equivalent to finding the largest root λ_* of the equation

$$\psi(\lambda) = 1. \tag{7.3}$$

Denote $\lambda_0 = 1 + \max_{1 \le i \le n} s^{(i)}$. Then $\psi(\lambda_0) \ge 0$ and therefore $\lambda_0 \le \lambda_*$. Consider the following process:

$$\lambda_{k+1} = \lambda_k - \frac{\psi(\lambda_k)}{\psi'(\lambda_k)}, \quad k \ge 0. \tag{7.4}$$

This is a standard Newton method for solving the equation (7.3), which can be also interpreted as a Newton method for minimization problem (7.2).

Lemma 7 *For any $k \geq 0$ we have $(\phi'(\lambda_k))^2 \leq n^7 \cdot \left(\frac{1}{16}\right)^k \phi''(\lambda_k)$.*

Proof:

Note that function ψ is decreasing and strictly convex. Therefore, for any $k \geq 0$ we have

$$\lambda_k \; < \; \lambda_{k+1} \; < \lambda_*, \quad \psi'(\lambda_k) \; < 0 \,, \quad \psi(\lambda_k) \; > \; 0.$$

Since $\psi(\lambda_k) \geq \psi(\lambda_{k+1}) + \psi'(\lambda_{k+1})(\lambda_k - \lambda_{k+1}) = \psi(\lambda_{k+1}) + \frac{\psi'(\lambda_{k+1})}{\psi'(\lambda_k)}\psi(\lambda_k)$, we obtain

$$1 \; \geq \; \tfrac{\psi(\lambda_{k+1})}{\psi(\lambda_k)} + \tfrac{\psi'(\lambda_{k+1})}{\psi'(\lambda_k)} \; \geq \; 2\sqrt{\tfrac{\psi(\lambda_{k+1})\psi'(\lambda_{k+1})}{\psi(\lambda_k)\psi'(\lambda_k)}}.$$

Thus, for any $k \geq 0$ we get

$$\phi''(\lambda_k) \cdot |\phi'(\lambda_k)| \; \leq \; \left(\tfrac{1}{4}\right)^k \phi''(\lambda_0) \cdot |\phi'(\lambda_0)|. \tag{7.5}$$

Further, in view of the choice of λ_0 we have

$$|\phi'(\lambda_0)| \; = \; \psi(\lambda_0) \; = \; \sum_{i=1}^{n} \tfrac{1}{\lambda_0 - s^{(i)}} - 1 \; < \; n - 1,$$

$$\phi''(\lambda_0) \; = \; \sum_{i=1}^{n} \tfrac{1}{(\lambda_0 - s^{(i)})^2} \; \leq \; n.$$

Finally, since $0 \leq \psi(\lambda_k) = \sum_{i=1}^{n} \tfrac{1}{\lambda_k - s^{(i)}} - 1$, we conclude that

$$\phi''(\lambda_k) \; = \; \sum_{i=1}^{n} \tfrac{1}{(\lambda_k - s^{(i)})^2} \; \geq \; \tfrac{1}{n}.$$

Using these bounds in (7.5), we obtain

$$\tfrac{1}{\phi''(\lambda_k)}(\phi'(\lambda_k))^2 \; \leq \; \left(\tfrac{1}{16}\right)^k \tfrac{(\phi''(\lambda_0))^2(\phi'(\lambda_0))^2}{(\phi''(\lambda_k))^3} \; \leq \; \left(\tfrac{1}{16}\right)^k \cdot n^7.$$

$\square$

Comparing statement of Lemma 7 with definition of $\mathcal{Q}(s)$, we conclude that the process (7.4) arrives at the region of quadratic convergence at most after

$$\left\lceil \tfrac{1}{4}(2 + 7\log_2 n) \right\rceil \tag{7.6}$$

iterations. Each of such iterations takes $O(n)$ arithmetic operations.

A similar technique can be used for finding the barrier projection in the cone of positive-semidefinite matrices:

$$\max_X \{\langle S, X \rangle + \ln \det X \; : \; \langle I, X \rangle = 1\}.$$

The most straightforward strategy consists in finding an eigenvalue decomposition of matrix S and solving the problem (7.1) with s being the spectrum of the matrix. In a more efficient strategy, we transform S in a tridiagonal form by an orthogonal transformation, compute its maximal eigenvalue and apply the Newton method to corresponding dual function.

Recent titles

CORE Discussion Papers

Recent titles

CORE Discussion Papers - continued

2008/48. David DE LA CROIX. Adult longevity and economic take-off: from Malthus to Ben-Porath.
2008/49. David DE LA CROIX and Gregory PONTHIERE. On the Golden Rule of capital accumulation under endogenous longevity.
2008/50. Jean J. GABSZEWICZ and Skerdilajda ZANAJ. Successive oligopolies and decreasing returns.
2008/51. Marie-Louise LEROUX, Pierre PESTIEAU and Grégory PONTHIERE. Optimal linear taxation under endogenous longevity.
2008/52. Yuri YATSENKO, Raouf BOUCEKKINE and Natali HRITONENKO. Estimating the dynamics of R&D-based growth models.
2008/53. Roland Iwan LUTTENS and Marie-Anne VALFORT. Voting for redistribution under desert-sensitive altruism.
2008/54. Sergei PEKARSKI. Budget deficits and inflation feedback.
2008/55. Raouf BOUCEKKINE, Jacek B. KRAWCZYK and Thomas VALLEE. Towards an understanding of tradeoffs between regional wealth, tightness of a common environmental constraint and the sharing rules.
2008/56. Santanu S. DEY. A note on the split rank of intersection cuts.
2008/57. Yu. NESTEROV. Primal-dual interior-point methods with asymmetric barriers.
2008/58. Marie-Louise LEROUX, Pierre PESTIEAU and Gregory PONTHIERE. Should we subsidize longevity?
2008/59. J. Roderick McCRORIE. The role of Skorokhod space in the development of the econometric analysis of time series.
2008/60. Yu. NESTEROV. Barrier subgradient method.

Books

Y. POCHET and L. WOLSEY (eds.) (2006), *Production planning by mixed integer programming.* New York, Springer-Verlag.
P. PESTIEAU (ed.) (2006), *The welfare state in the European Union: economic and social perspectives.* Oxford, Oxford University Press.
H. TULKENS (ed.) (2006), *Public goods, environmental externalities and fiscal competition.* New York, Springer-Verlag.
V. GINSBURGH and D. THROSBY (eds.) (2006), *Handbook of the economics of art and culture.* Amsterdam, Elsevier.
J. GABSZEWICZ (ed.) (2006), La différenciation des produits. Paris, La découverte.
L. BAUWENS, W. POHLMEIER and D. VEREDAS (eds.) (2008), *High frequency financial econometrics: recent developments.* Heidelberg, Physica-Verlag.
P. VAN HENTENRYCKE and L. WOLSEY (eds.) (2007), *Integration of AI and OR techniques in constraint programming for combinatorial optimization problems.* Berlin, Springer.

CORE Lecture Series

C. GOURIÉROUX and A. MONFORT (1995), Simulation Based Econometric Methods.
A. RUBINSTEIN (1996), Lectures on Modeling Bounded Rationality.
J. RENEGAR (1999), A Mathematical View of Interior-Point Methods in Convex Optimization.
B.D. BERNHEIM and M.D. WHINSTON (1999), Anticompetitive Exclusion and Foreclosure Through Vertical Agreements.
D. BIENSTOCK (2001), Potential function methods for approximately solving linear programming problems: theory and practice.
R. AMIR (2002), Supermodularity and complementarity in economics.
R. WEISMANTEL (2006), Lectures on mixed nonlinear programming.